AD_____

Award Number: DAMD17-01-1-0608

TITLE: Determination of a Unique Pattern of Gene Expression in Node Positive Breast Cancer Using Serial Analysis of Gene Expression (SAGE)

PRINCIPAL INVESTIGATOR: Adam M. Brufsky, Ph.D.

CONTRACTING ORGANIZATION: University of Pittsburgh
Pittsburgh, Pennsylvania 15260

REPORT DATE: June 2002

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2002	Annual (1 Jun 01 - 30 Jun 02)

4. TITLE AND SUBTITLE
Determination of a Unique Pattern of Gene Expression in Node Positive Breast Cancer Using Serial Analysis of Gene Expression (SAGE)

5. FUNDING NUMBERS
DAMD17-01-1-0608

6. AUTHOR(S)
Adam M. Brufsky, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Pittsburgh
Pittsburgh, Pennsylvania 15260

E-Mail: brufskyam@msx.upmc.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

No abstract provided.

14. SUBJECT TERMS
No subject terms provided.

15. NUMBER OF PAGES
9

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Progress Report
DOD Concept Award
PI: Adam Brufsky, MD, PhD

Individual breast cancer cells in a growing primary tumor are genetically unstable and undergo somatic mutation at an accelerated rate. Primary breast tumors are therefore composed of genetically heterogeneous clones of cells. Individual clones of cells, through activation and expression of normally suppressed genes, likely acquire the capacity for lymphogenous and hematogenous spread at a very early stage of primary tumor development. These clones likely represent only a small fraction of the total tumor volume. We hypothesize that a pattern of genetic expression (a "signature") for lymphogenous spread exists in a small fraction of cells of a primary breast cancer. In addition, we hypothesize that this signature will be constant between primary tumors from different women. This genetic signature should be detectable by sensitive genetic differential display methods, and can be exploited to determine the lymph node status of a primary breast cancer. A corollary to this hypothesis is that a lymph node involved by metastatic breast cancer should represent an in vivo amplification of clones of cells of a primary breast cancer. The lymph node should have a pattern of genetic expression differing from the primary tumor. In addition, the lymph node genetic expression signature should contain genes associated with increased aggressiveness.

In preliminary work, we have used cDNA microarrays to analyze mRNA expression patterns in a primary tumor-lymph node pairs from tissue flash frozen in liquid nitrogen within five minutes of removal at surgery. In these experiments, we were careful to use a large tumor from which all normal breast tissue had been trimmed. To reduce lymphoid cell contamination, the matched lymph node used for mRNA extraction was completely replaced by tumor. mRNA was converted to 33P-labelled cDNA and used to sequentially probe Research Genetics cDNA nylon filter microarrays containing approximately 30,000 cDNA clones. Differences in mRNA expression between the lymph node and the tumor were analyzed by computer software provided by Research Genetics. Alternatively, mRNA was converted to fluorescently labeled cDNA and used to screen Incyte GEM-1 glass based microarrays. A total of five paired primary tumor/lymph node samples were assayed by these methods.

Experiment 1 (Research Genetics cDNA array): A total of 249 genes were > 2-fold overexpressed in the lymph node compared to the paired tumor sample and several representative genes are shown in Table 1.1. Overexpression of several of these genes, including thymosin beta 4 and AIB1, has been associated with increased aggressiveness of primary breast cancer. In contrast only a single gene was found to be more highly expressed in the tumor sample relative to the lymph node.

Table 1.1.

Differential gene expressic (LN vs TUM)	Gene ID
Higher in LN vs TUM	
4.243519	Zinc finger protein 9 (a cellular retroviral nucleic acid binding protein)
3.921545	"Human thymosin beta-4 mRNA, complete cds"
3.617571	"Human guanine nucleotide-binding protein G-s, alpha subunit mRNA, partial cds"
3.598648	EST
3.568079	"Human guanine nucleotide-binding protein G-s, alpha subunit mRNA, partial cds"

3.498626	Zinc finger protein 43 (HTF6)
3.432767	"Human guanine nucleotide-binding protein G-s, alpha subunit mRNA, partial cds"
3.307328	H.sapiens mRNA for elongation factor-1-gamma
3.286589	"Human guanine nucleotide-binding protein G-s, alpha subunit mRNA, partial cds"
3.239934	ESTs
Lower in LN vs TUM	
-1.78794	ESTs
-1.78863	AF-9 PROTEIN
-1.78882	EST
-1.79336	ESTs
-1.79832	H.sapiens mRNA for uridine phosphorylase
-1.82785	ESTs
-1.91391	"ESTs, Highly similar to HYPOTHETICAL 36.7 KD PROTEIN C2F7.02C IN CHROMOSOME I [Schizosaccharomyces pombe]"
-1.93279	ESTs
-1.97063	"ESTs, Moderately similar to RAS-RELATED PROTEIN RAB-28 [H.sapiens]"
-2.1818	Human mRNA for alanine aminotransferase
-1.78794	ESTs

Experiment 2 (Research Genetics cDNA array): A total of 24 genes were > 2-fold overexpressed in the lymph node compared to the paired tumor sample and several representatives are shown in Table 1.2. Furthermore, 155 genes were found to be more highly expressed in the tumor sample relative to the lymph node. The differentially expressed genes identified in these samples appear reflect samples contamination with normal tissue rather than markers of matastatic disease. For example, many extracellular matrix components are highly expressed in the primary tumor, whereas lymphatic markers are highly expressed in the lymph node. Alternatively these data may reflect the fact that tumor cells to some extent adopt the gene expression profiles of their surroundings.

Table 1.2.

Differential gene expressio (LN vs TUM)	Gene ID
Higher in LN vs TUM	
4.14731	Immunoglobulin lambda light chain
4.435978	ESTs
4.590368	ESTs
4.901573	H.sapiens mRNA for elongation factor-1-gamma
5.093006	COMPLEMENT FACTOR H-LIKE PROTEIN DOWN16 PRECURSOR
5.363208	EST
6.875204	Human Ig germline H-chain G-E-A region B: gamma-2 constant region, 3' end
7.266351	Human Ig germline H-chain G-E-A region B: gamma-2 constant region, 3' end
7.656507	Human AMP deaminase (AMPD2) mRNA
11.64051	Human rearranged immunoglobulin lambda light chain mRNA
Lower in LN vs TUM	
-18.158	EST
-12.9861	Collagen, type I, alpha-2
-10.8633	Homo sapiens mRNA for osteoblast specific factor 2 (OSF-2os)
-9.00068	Alpha-1 type 3 collagen
-8.55227	Ferritin, light polypeptide
-8.37082	Fibronectin 1
-7.40909	Fibronectin 1
-5.08964	Human SWI/SNF complex 60 KDa subunit (BAF60b) mRNA, complete cds

-4.30908	Collagen, type I, alpha-2
-3.7871	Alpha-1 type 3 collagen

Experiment 3 (Research Genetics cDNA array): A total of 5 genes were > 2-fold overexpressed in the lymph node compared to the paired tumor sample. Furthermore, 23 genes were more highly expressed in the tumor sample relative to the lymph node. Some of the differentially expressed genes identified in these samples again may reflect slight tumor sample contamination with normal tissue in that a small number of extracellular matrix components are highly expressed in the primary tumor, whereas lymphatic markers are highly expressed in the lymph node.

Table 1.3.

Differential gene expressio (LN vs TUM)	Gene ID
Higher in LN vs TUM	
2.540519	ESTs
2.418438	ESTs
2.15137	ESTs, Weakly similar to T-LYMPHOCYTE MATURATION-ASSOCIATED PROTEIN [H.sapiens]
2.091487	Homo sapiens mRNA for klotho, complete cds
2.031324	ESTs, Highly similar to similar to mouse CC1. [H.sapiens]
1.995861	ESTs, Moderately similar to fibrosin [M.musculus]
1.956819	ER LUMEN PROTEIN RETAINING RECEPTOR 2
1.911267	Homo sapiens inositol polyphosphate 4-phosphatase type II-alpha mRNA, complete cds
1.909991	ESTs
1.896783	Cardiac gap junction protein
Lower in LN vs TUM	
-5.1763	Human mercurial-insensitive water channel mRNA, form 2, complete cds
-2.61545	ESTs
-2.49583	Human glutamate receptor 2 (HBGR2) mRNA, complete cds
-2.36191	Human dihydrolipoamide dehydrogenase mRNA, complete cds
-2.33229	Collagen, type I, alpha-2
-2.2355	Lactotransferrin
-2.1122	Connective tissue growth factor
-2.10446	unknown EST
-2.02818	Alpha-1 type 3 collagen
-2.04594	ESTs, Highly similar to OSH1 PROTEIN [Saccharomyces cerevisiae]

Experiment 4 (Incyte cDNA array): A total of 73 genes were > 2-fold overexpressed in the lymph node compared to the paired tumor sample. Furthermore, 98 genes were more highly expressed in the tumor sample relative to the lymph node. Interestingly, the most highly overexpressed gene in lymph node is estrogen receptor (ER) 1, suggesting that clonal expansion of ER positive tumor cells may have occurred during metastasis. This is supported by the high level of expression of many estrogen-responsive genes including LIV-1 and lactotransferrin. As in previous experiments, extracellular matrix gene expression is higher in the primary tumor, whilst lymphatic markers are more highly expressed in the lymph node.

Table 1.4.

Differential gene expressio (LN vs TUM)	Gene ID
Higher in LN vs TUM	
21.8	estrogen receptor 1 {Incyte PD: 4116386}
11.6	N-acetyltransferase 1 (arylamine N-acetyltransferase) {Incyte PD: 2613155}

8.7	LIV-1 protein, estrogen regulated {Incyte PD: 1402273}
8.7	S100 calcium-binding protein P {Incyte PD: 2060823}
8	carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross reacting antigen) {Incyte PD: 2060355}
7.5	Human secretory protein (P1.B) mRNA, complete cds {Incyte PD: 2242817}
6.9	carcinoembryonic antigen-related cell adhesion molecule 8 {Incyte PD: 172588}
6.6	coxsackie virus and adenovirus receptor {Incyte PD: 1282194}
6.2	tetraspan 1 {Incyte PD: 2989680}
6.1	N-acetyltransferase 2 (arylamine N-acetyltransferase) {Incyte PD: 86365}
Lower in LN vs TUM	
-14	S100 calcium-binding protein A9 (calgranulin B) {Incyte PD: 4283945}
-10.7	Homo sapiens clone 24636 mRNA sequence {Incyte PD: 1552481}
-10.3	phosphoinositide-3-kinase, class 3 {Incyte PD: 9140}
-8.2	ESTs {Incyte PD: 2816379}
-7.2	apolipoprotein D {Incyte PD: 551403}
-6.3	collagen, type I, alpha 1 {Incyte PD: 782235}
-5.3	cysteine-rich, angiogenic inducer, 61 {Incyte PD: 1514989}
-5	epidermal growth factor receptor (avian erythroblastic leukemia viral (v-erb-b) oncogene homolog) {Incyte PD: 179598}
-4.5	peripheral myelin protein 2 {Incyte PD: 2963131}

Experiment 5 (Incyte cDNA array): A total of 30 genes were > 2-fold overexpressed in the lymph node compared to the paired tumor sample. Furthermore, 174 genes were more highly expressed in the tumor sample relative to the lymph node. In contrast to experiment 4 (above), the most highly overexpressed gene in the primary tumor is estrogen receptor (ER) 1. As above, extracellular matrix gene expression is higher in the primary tumor, whilst lymphatic markers are more highly expressed in the lymph node.

Table 1.5.

Differential gene expressio (LN vs TUM)	Gene ID
Higher in LN vs TUM	
7.4	membrane-spanning 4-domains, subfamily A, member 2 {Incyte PD: 22882}
6.2	ESTs {Incyte PD: 2313368}
5.8	lymphotoxin beta (TNF superfamily, member 3) {Incyte PD: 508631}
4.7	selectin L (lymphocyte adhesion molecule 1) {Incyte PD: 1876370}
4.5	CD79B antigen (immunoglobulin-associated beta) {Incyte PD: 1646010}
4.5	complement component (3d/Epstein Barr virus) receptor 2 {Incyte PD: 3055203}
4.5	Homo sapiens clone 24636 mRNA sequence {Incyte PD: 1552481}
4.2	Fc fragment of IgE, low affinity II, receptor for (CD23A) {Incyte PD: 4411157}
3.3	CD3D antigen, delta polypeptide (TiT3 complex) {Incyte PD: 3297914}
3.3	immunoglobulin gamma 3 (Gm marker) {Incyte PD: 3553751}
Lower in LN vs TUM	
-23.4	osteoblast specific factor 2 (fasciclin I-like) {Incyte PD: 1994715}
-12.3	estrogen receptor 1 {Incyte PD: 4116386}
-10.8	collagen, type I, alpha 1 {Incyte PD: 782235}
-10.8	fibronectin 1 {Incyte PD: 3553729}
-9.3	Human secretory protein (P1.B) mRNA, complete cds {Incyte PD: 2242817}
-8.9	lumican {Incyte PD: 1228124}
-8.8	Human Ig J chain gene {Incyte PD: 1001933}
-7.3	v-jun avian sarcoma virus 17 oncogene homolog {Incyte PD: 1969563}
-6.5	S100 calcium-binding protein P {Incyte PD: 2060823}

-6.4	collagen, type XI, alpha 1 {Incyte PD: 3598222}

Summary of microarray experiments:

Microarray analysis to compare paired samples of primary breast tumor and metastatic axillary lymph node reveal the differential expression of a numner of genes that have previously been associated with metastasis and tumor progression. Experiments 4 and 5 are particularly interesting. The high lymph node overexpression (relative to the primary tumor) of ER in experiment 4 suggests that clonal expansion of ER positive tumor cells may have occurred during metastasis. In contrast low lymph node ER expression relative to the primary tumor in experiment 5 suggests a dilution of ER positive tumor cells via clonal expansion of ER negative cells. Interestingly, a large number of other genes are differentially expressed in a similar pattern between these two samples, suggesting that these are estrogen responsive. The fact that extracellular matrix gene expression is higher in both primary tumors, whilst lymphatic markers are more highly expressed in the lymph nodes (experiments 4 and 5) suggests that these ER-specific differences in genes expression are not just the result of contamination by normal tissue. A summery of similarities and differences in gene expression between the samples assayed in experiments 4 and 5 is contained in the attached spreadsheet. Further analyses, using TaqMan RTPCR and histological methods are under way to confirm these potentially interesting findings.

Experiment 6 (SAGE analysis):

We have begun SAGE analysis to compare differences in gene expression between node positive and node negative primary breast tumors. These experiments are in the data gathering stage.

SAGE is an unbiased sampling method in which the steady state concentration of potentially every distinct RNA molecule in a given cell or tissue type is determined. This is a distinct advantage of SAGE over microarrays in that it provides an absolute, rather than relative, measure of gene expression. The expression level of a specific gene can thus be compared to the levels of every other gene in that (and any other) sample and, therefore, SAGE generates immortal data that represents absolute levels of expression. This kind of detail is extremely important when one is trying to construct meaningful biological models of complex gene expression data sets. Furthermore, whole series of SAGE libraries, even those generated by different investigators, may be directly compared. This is in direct contrast to cDNA array approaches, which always demand the simultaneous analysis of a reference sample. Also, unlike microarrays, SAGE does not require any prior knowledge of the genes of interest and is able to interrogate the entire genome (or "transcriptome") in a completely unbiased fashion. SAGE also allows the discovery and analysis of novel genes.

We are currently in the process of analyzing our SAGE data. We plan to apply for NIH R01 or R03 funding for a clinical trial to predict lymph node positivity from core breast biopsies using the gene expression pattern determined from our studies.

Additional Studies performed under this grant:

We performed additional microarray studies on LN(+) and LN(-) breast cancer using Affymetrix microarrays and submitted an abstract to the 2003 meeting of the American Society of Clinical Oncology. Gene expression profiling has the potential to classify primary breast tumors into prognostic groups. A 70-gene prognosis profile (NEJM 2002; 347: 1999) was recently developed by a supervised

classification of gene expression data from young women with lymph node negative breast cancer, and was used to stratify a larger group of 295 women into good and poor prognosis groups independent of lymph node status. Lymph node status, however, remains the best clinical predictor of patient outcome. We believe that a prognostic profile of breast cancer based on a gene expression profile unique to lymph node positive breast cancer would have clinical utility. In addition, prior work with gene prognosis profiling, including the above study, has been performed on a set of primary tumors heterogeneous for hormone receptor and Her2 Neu status. To address these issues, we have developed a gene expression profile of ER (-), PR (-), Her2 (-), lymph node positive breast cancer. RNA was extracted from 5 LN(+) and 5 LN (-) primary breast tumors and gene expression was analyzed in duplicate by Affymetrix U133A microarrays containing 22,284 genes. Relative gene expression produced a preliminary expression profile of 18 up-regulated genes (>1.5 fold, $p< 0.01$ by t-test) and 40 down-regulated genes (>1.5 fold, $p<0.01$) unique to lymph node positive breast cancer. Up-regulated genes and ESTs include those related to v-maf, IFNGR1, RAD23A, MAFF. Down-regulated genes and ESTs include those related to Enolase 3, TFDP2, TP63 and Cyclin E2. Data from a larger validation set will be presented. This profile differs significantly from poor prognostic profiles found in hormone receptor mixed sample sets.